УРАЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ
ЛЕСОТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ
УГЛТУ

А.Г. Гороховский
А.В. Мялицин

НАУЧНЫЕ ИССЛЕДОВАНИЯ В ДЕРЕВООБРАБОТКЕ

Екатеринбург
2012

МИНОБРНАУКИ РОССИИ

ФГБОУ ВПО «УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ЛЕСОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра древесиноведения и специальной обработки древесины

А.Г. Гороховский

А.В. Мялицин

НАУЧНЫЕ ИССЛЕДОВАНИЯ В ДЕРЕВООБРАБОТКЕ

Методические указания
для лабораторных работ по курсу
«Методы и средства научных исследований»
для студентов направления 250300 «Технология и оборудование
лесозаготовительных и деревообрабатывающих производств»
очной формы обучения

Екатеринбург
2012

Печатается по рекомендации методической комиссии факультета механической технологии древесины. Протокол № 2 от 05.09.2011 г.

Рецензент – канд. тех. наук, доцент, зав. кафедрой МОД О.Н. Чернышев

Редактор К.В. Корнева
Оператор компьютерной верстки Е.В. Карпова

Подписано в печать 30.03.2012		Поз. 78
Плоская печать	Формат 60х84 1/16	Тираж 50 экз.
Заказ №	Печ. л. 2,09	Цена 10 р. 72. коп.

Редакционно-издательский отдел УГЛТУ
Отдел оперативной полиграфии УГЛТУ

ВВЕДЕНИЕ

Современное производство требует от специалиста принятия квалифицированных инженерных решений при проектировании, изготовлении и эксплуатации технологического оборудования. Умение проводить научные исследования становится необходимостью, так как часто лишь с их помощью удаётся учесть особенности конкретных условий производства и выявить резервы повышения его эффективности.

Подготовка будущих специалистов должна в этой связи включать не только изучение основ техники и технологии, но и методологии проведения научно-исследовательских и опытно-конструкторских работ. Знание физики процесса в совокупности с научно обоснованным и грамотно поставленным экспериментом позволяют исследователю иметь чёткое представление о сущности протекающих в рассматриваемой системе процессов, выявлять факторы и условия, влияющие на их ход, определять направление движения к оптимальным структуре, конструктивным и режимным параметрам технологических процессов и оборудования.

Сложность задач, решаемых при проведении научных исследований, обусловливает применение компьютерных технологий, поэтому для современного исследователя важно умение использовать различные пакеты прикладных программ, позволяющих проводить обработку экспериментальных данных и моделирования процессов.

Выводы, полученные в результате проведения исследования, должны иметь практическое применение в организации технологического процесса или в конструкции оборудования. Такие выводы могут быть как организационно-технического характера, так и имеющие отношение к изобретательской деятельности.

Исходя из выше изложенного, в методических указаниях излагаются основы проведения теоретических и экспериментальных исследований, методика обработки экспериментальных данных, в том числе с применением физико-математического пакета Mathcad и офисного приложения Microsoft Exel.

Приложения содержат основные команды, используемые при математической обработке экспериментальных данных в пакетах Excel и Mathcad, таблицы определения критериев Стъюдента, Кохрена, Фишера, Пирсона.

ЛАБОРАТОРНАЯ РАБОТА № 1

ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК СЛУЧАЙНЫХ ВЕЛИЧИН

Для объективной оценки полученных результатов исследования необходима их математическая обработка.

Множество значений случайной величины, полученных в результате эксперимента или наблюдений над объектом исследования, представляет собой статистическую совокупность. Статистическая совокупность, содержащая в себе все возможные значения случайной величины, называется генеральной статистической совокупностью. Выборочной статистической совокупностью или выборкой называют совокупность, в которой содержится только некоторая часть элементов генеральной совокупности. Число опытов, содержащихся в выборке, называют объемом выборки.

Обработка экспериментальных данных проводится с помощью методов математической статистики.

Математическая обработка включает расчет как минимум следующих статистических величин:

• среднее арифметическое – сумма значений, полученных по результатам испытания выборки, деленная на их объем;

• дисперсия – средний квадрат отклонения отдельных результатов наблюдений от среднего значения случайной величины:

• среднеквадратичное отклонение единичного результата – квадратный корень из дисперсии;

• стандартное отклонение средней арифметической или ошибка средней арифметической из всех n повторений;

• доверительная ошибка оценки измеряемой величины.

Кроме того, при изучении исследователем влияния каких-либо факторов на параметр технологического процесса необходимо также устанавливать коэффициент корреляции и функциональную зависимость между ними.

Указанные выше статистические величины рассчитываются по приведенным ниже формулам.

Среднее арифметическое:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \qquad (1)$$

где n – число наблюдений;

x_i – значение единичного измерения величины.

Дисперсия:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2. \tag{2}$$

Среднее квадратичное отклонение:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}. \tag{3}$$

Стандартное отклонение или средняя квадратическая ошибка среднего арифметического:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}, \text{ при } n > 30,$$
$$s_{\bar{x}} = \frac{s}{\sqrt{n-1}}, \text{ при } n < 30. \tag{4}$$

Коэффициент вариации:

$$V = 100\frac{s}{\bar{x}}. \tag{5}$$

Показатель точности среднего значения:

$$\xi = \frac{s_{\bar{x}}}{\bar{x}}100. \tag{6}$$

Доверительный интервал.

Истинное значение измеряемой величины с наперед заданной доверительной вероятностью (Р) должно лежать в пределах доверительного интервала $\bar{x} \pm \Delta$.

$$\Delta = \frac{ts}{\sqrt{n}}. \tag{7}$$

Для определения доверительного интервала результата используется критерий Стьюдента t (Р; f). Критерий t (Р; f) берется из таблицы (прил. 3) в зависимости от уровня значимости q = 1 – Р и числа степеней свободы f = n – 1. Для практических целей в области деревопереработки вполне достаточным является значение q = 0,05 %.

ЗАДАНИЯ К ЛАБОРАТОРНОЙ РАБОТЕ № 1

1. Вычислить выборочное квадратическое отклонение, коэффициент вариации, среднюю квадратическую ошибку среднего арифметического, показатель точности. Сделать вывод о надежности полученных результатов.
2. Определить доверительный интервал для математического ожидания.

3. Задавшись по своему усмотрению каким-либо другим доверительным интервалом, определить необходимый объем выборки.

ЛАБОРАТОРНАЯ РАБОТА № 2

МЕТОД ГРУППИРОВАНИЯ ДАННЫХ

Для построения гистограммы статистического распределения результатов наблюдений, прежде всего необходимо произвести их группирование, то есть разделение ряда данных от наименьшего x_{min} до наибольшего x_{max} на r интервалов.

Предварительное количество интервалов r из объема выборки N вычисляется по формуле:

$$r = 1 + 3{,}2\lg N. \tag{8}$$

Количество интервалов должно быть целым числом, т.е. полученное значение следует округлить до целого значения.

Ширину интервала Δx_i(i = 1, 2, ..., r) выбирают постоянной для всего ряда данных, т.е. $\Delta x_i = \frac{x_{max} - x_{min}}{r}$. После этого подсчитывают числа m_i, равные числу результатов, попадающих в каждый i-й интервал, т.е меньших или равных его правой и больших его левой границы. Отношения $P_i^* = \frac{m}{N}$ представляют собой статистические оценки вероятностей попадания результата наблюдений в i-й интервал. Если частоты попадания результата наблюдений разделить на длину интервала, то получим значение $f_i^* = \frac{P_i^*}{\Delta x_i}$, являющееся оценкой средней плотности распределения в интервале Δx_i. Отложив вдоль оси результатов наблюдений, как показано на рисунке 1, интервалы Δx_i в порядке возрастания индекса i и построив на каждом интервале прямоугольник с высотой, равной f_i^*, получим график, называемый **гистограммой** статистического распределения.

Наиболее наглядной формой представления выборок является гистограмма.

Если гистограмму модифицировать следующим образом – соединить отрезками прямых середины горизонтальных отрезков, то полученная ломанная является графиком непрерывной функции и называется **полигоном частот**.

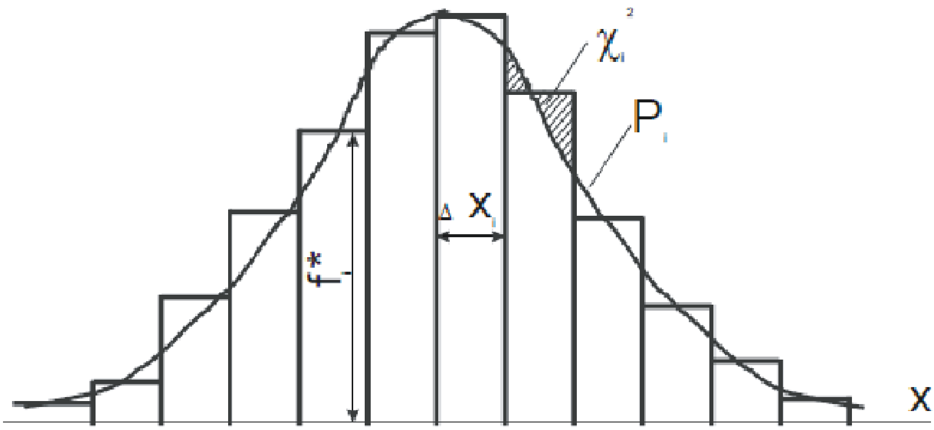


Рис. 1. Гистограмма статистического распределения

Порядок построения гистограммы в Excel:

1. Намечаем исследуемый показатель.
2. Проводим измерения. Должно быть не менее 30–50 данных, оптимально – около 100.
3. Вводим единицу измерений. Единица измерений равна точности, с которой проводились измерения.
4. Находим минимальное и максимальное значения выборки. Минимальное и максимальное значения выборки находим с помощью статистических функций МИН и МАКС.
5. Находим размах выборки R как разность между максимальным и минимальным значениями выборки.
6. Определяем предварительное количество интервалов $K_{пр}$ из объёма выборки N.
7. Определяем ширину интервала по формуле $h = R/ K_{пр}$ с округлением до единицы измерения.
8. Вводим номера интервалов с 1 примерно до 25.
9. Рассчитываем границы и середины интервалов. Рассчитываем нижнюю границу первого интервала по формуле:

$$\frac{X_{min}\text{-ед.изм.}}{2}. \tag{9}$$

Определяем нижнюю границу второго интервала, прибавляя к нижней границе первого интервала значение шага. Рассчитываем верхнюю границу первого интервала, прибавляя к его нижней границе значение шага.

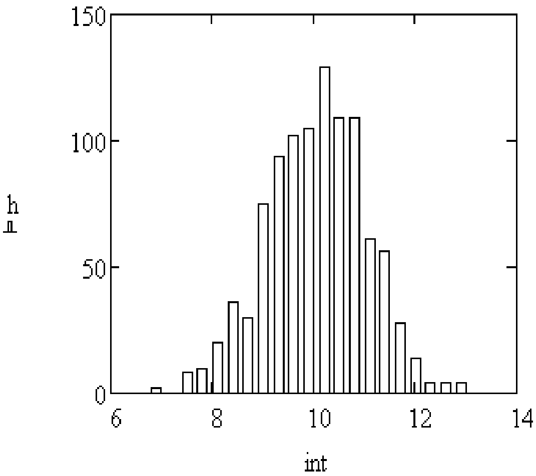
10. Подсчитываем частоты появления результатов измерений в интервалах. Рассчитываем частоту для первого интервала при помощи статистической функции СЧЁТЕСЛИ. Функция СЧЁТЕСЛИ подсчитывает количество непустых ячеек в указанном диапазоне, удовлетворяющих заданному условию.
11. Строим гистограмму распределения.

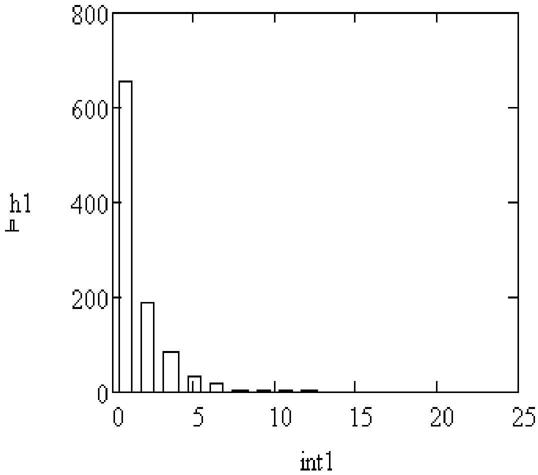
Результаты расчётов показаны на рисунке 2.

№	Коэф. деформ.	Ед. изм.=	0,1			
1	0,9	Xмин =	0,1			
2	0,6	Xмах =	1,8			
3	0,5	R =	1,7			
4	0,6	Кпредв =	10			
5	0,7	h =	0,2			
6	0,8					
7	1	№ инт.	Ниж. гр.	Верх. гр.	Середина	Частота f
8	1,4	1	0,05	0,25	0,15	2
9	1,1	2	0,25	0,45	0,35	8
10	1,5	3	0,45	0,65	0,55	13
11	1,5	4	0,65	0,85	0,75	15
12	0,1	5	0,85	1,05	0,95	20
13	0,8	6	1,05	1,25	1,15	17

Рис. 2. Расчёт данных для построения гистограммы в Excel

Пример построения гистограмм приведен на рисунке 3.





А Б

Рис. 3. Построение гистограмм в Mathcad:
А – для нормального распределения величин;
Б – для логарифмического распределения величин

Для определения типа закономерности эмпирического распределения оно приближенно описывается подходящим теоретическим (вероятностным) распределением, форму кривой которого называют формат распределения. В тех случаях, когда форма распределения анализируется на ее близость к нормальной форме, расхождение между ними оценивается показателями асимметрии и эксцесса.

Коэффициент асимметрии

Показатели асимметрии оценивают смещение ряда распределения влево или вправо по отношению к оси симметрии нормального распределения:

$$A = \frac{1}{ns^3}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3. \qquad (10)$$

Установлена следующая оценочная шкала асимметричности:

$|A| \le 0{,}25$ – асимметрия незначительная;

$0{,}25 < |A| \le 0{,}5$ – асимметрия заметная (умеренная);

$|A| > 0{,}5$ – асимметрия существенная.

Коэффициент эксцесса

Показатель эксцесса характеризует крутизну кривой распределения – ее заостренность или пологость по сравнению с нормальной кривой.

$$E = \frac{1}{ns^4}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4 - 3. \qquad (11)$$

Как правило, коэффициент эксцесса вычисляется только для симметричных или близких к ним распределений. Это объясняется тем, что за базу сравнения принята кривая нормального распределения, являющаяся симметричной. Относительно вершины нормальной кривой и определяется выпад вверх или вниз вершины теоретической кривой эмпирического распределения. При этом:

– если $E > 0$, то вершина кривой распределения располагается выше вершины нормальной кривой, а форма кривой является более островершинной, чем нормальная. Это говорит о скоплении значений признака в центральной зоне ряда распределения, т.е. о преимущественном появлении в данных значений, близких к средним;

– если $E < 0$, то вершина кривой распределения лежит ниже вершины нормальной кривой, а форма кривой более пологая по сравнению с нормальной. Это означает, что значения признака не концентрируются в центральной части ряда, а достаточно равномерно рассеяны по всему диапазону от x_{max} до x_{min}.

Для нормального распределения $E = 0$, поэтому чем больше абсолютная величина $|E|$, тем существеннее распределение отличается от нормального. В частности, большая отрицательная величина E означает преобладание у признака крайних значений, причем одновременно и более низких, и более высоких. При этом в центральной части распределения может образоваться «впадина», превращающая распределение в двухвершинное (U-образной формы), что является индикатором неоднородности совокупности.

Средние квадратические отклонения для ассиметрии и эксцесса:

$$\sigma_A = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}};$$
$$\sigma_E = \sqrt{\frac{24(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}. \qquad (12)$$

Если одна из характеристик A или E по абсолютной величине существенно, в 2–3 раза, превосходит соответствующее среднее квадратическое отклонение, то следует усомниться в нормальности распределения и провести более тщательную проверку с помощью критерия Пирсона или Колмогорова.

ЗАДАНИЯ К ЛАБОРАТОРНОЙ РАБОТЕ № 2

1. Определить среднее арифметическое и дисперсию выборки.
2. Рассчитать выборочные показатели асимметрии и эксцесса.
3. Рассчитать квадратические отклонения для асимметрии и эксцесса.
4. Построить гистограмму и полигон частот выборки.

ЛАБОРАТОРНАЯ РАБОТА № 3

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

3.1. ОТБРАСЫВАНИЕ ГРУБЫХ НАБЛЮДЕНИЙ

Грубые наблюдения (промахи) необходимо из выборки исключить. Для проверки предположения, является сомнительный результат x_i промахом или нет, его временно исключают из выборки и по оставшимся наблюдениям определяют среднее арифметическое $\bar{x}$ и оценку дисперсии s^2. Затем определяют расчетный критерий Стьюдента:

$$t_{расч} = \frac{|x_i - \bar{x}|}{s}. \qquad (13)$$

Из таблиц распределения Стьюдента по выбранному уровню значимости q и числу степеней свободы f находят табличное значение t – критерия $t_{таб}$. Если $t_{расч} > t_{таб}$, то подозреваемый результат является промахом и должен быть исключен из выборки.

Иногда сомнение вызывают одновременно два или более элементов выборки, их также из выборки исключают, рассчитывают значения $\bar{x}$ и s^2. Затем решают вопрос об исключении элемента, значение которого ближе к среднему арифметическому с использованием изложенного метода выше. Если это наблюдение окажется промахом, оставшиеся сомнительные наблюдения будут также промахами. Однако если менее сомнительный вариант не окажется промахом, его присоединяют к выборке, исследуют следующий сомнительный вариант и т.д.

3.2. ПРОВЕРКА ГИПОТЕЗЫ ОБ ОДНОРОДНОСТИ ДВУХ ДИСПЕРСИЙ

Дисперсии двух совокупностей объемами n_1 и n_2, подчиняющиеся нормальному (логарифмически нормальному) закону распределения, сравнивают с помощью двустороннего критерия Фишера F. Для этого рассчитывают дисперсионное отношение F по формуле:

$$F = \frac{s_1^2}{s_2^2} \text{ при } s_1^2 > s_2^2, \tag{14}$$

где s_1^2, s_2^2 – выборочные дисперсии.

Дисперсионное отношение F сопоставляют с критическим значением $F_{таб}$ для заданного уровня значимости q и чисел степеней свободы $f_1 = n_1 - 1, f_2 = n_2 - 1$, где f_1 – число степеней свободы для большей дисперсии. В случае соблюдения условия $F \le F_{таб}$ принимают гипотезу об однородности дисперсий. В противном случае нулевая гипотеза отвергается. Критические значения критерия Фишера приведены в приложении 5.

3.3. ПРОВЕРКА ОДНОРОДНОСТИ НЕСКОЛЬКИХ ДИСПЕРСИЙ, НАЙДЕННЫХ ПО ВЫБОРКАМ ОДИНАКОВОГО ОБЪЕМА

Использование F – критерия Фишера – при числе более двух неэффективно, так как при этом в оценке участвуют только наибольшая и наименьшая дисперсии. Критерий Кохрена пригоден для случаев, когда число повторных опытов n во всех точках плана одинаково. Критерий Кохрена – это отношение максимальной дисперсии к сумме всех дисперсий:

$$G = \frac{s_{max}^2}{\sum\limits_{i=1}^{m} s_m^2}, \tag{15}$$

где m – количество выборочных дисперсий.

Затем по выбранному уровню значимости q, числу степеней свободы выборок $m_1 = n - 1$ и по количеству выборок m_2 из таблицы находят величину

$G_{таб}$. Если $G < G_{таб}$, то можно принять гипотезу об однородности дисперсий. В противном случае она отвергается.

3.4. ПРОВЕРКА ОДНОРОДНОСТИ НЕСКОЛЬКИХ ДИСПЕРСИЙ, НАЙДЕННЫХ ПО ВЫБОРКАМ РАЗЛИЧНОГО ОБЪЕМА

Экспериментаторы часто планируют получение выборок одинакового объема, однако, если в опытах обнаруживаются промахи, то после их исключения объемы выборок оказываются различными. Пусть проверяется однородность некоторого числа m дисперсий: $s_1^2, s_2^2, s_3^2, \ldots, s_m^2$. Теперь эти дисперсии найдены по выборкам различного объема – соответственно, $n_1, n_2, n_3, \ldots, n_m$. В этом случае используют критерий Бартлетта. Предварительно вычисляют величину s_y^2, представляющую собой среднее взвешенное значение дисперсий, взятое с учетом числа степеней свободы:

$$s_y^2 = \frac{\sum\limits_{i=1}^{m} f_i s_i^2}{f}, \tag{16}$$

где $f = \sum\limits_{i=1}^{m} f_i$ – числа степеней свободы соответствующих дисперсий.

$$f_i = n_i - 1. \tag{17}$$

Далее рассчитывают величину B = V/C. Здесь V и C равны:

$$\begin{gathered} V = 2{,}303\left(f \lg s_y^2 - \sum_{i=1}^{m} f_i \lg s_i^2 \right); \\ C = 1 + \frac{1}{3(m-1)}\left(\sum_{i=1}^{m} \frac{1}{f_i} - \frac{1}{f} \right). \end{gathered} \tag{18}$$

Затем из приложения 6 при уровне значимости q и числе степеней свободы k = m – 1 отыскивают значение $\chi^2_{таб}$. Гипотеза об однородности дисперсий принимается, если $B \le \chi^2_{таб}$. В данной проверке требуется, чтобы объем каждой выборки был не менее четырех. Применение критерия Бартлетта является трудоемким. Кроме того, следует иметь в виду, что он весьма чувствителен к отклонениям от нормальности распределения.

3.5. ПРОВЕРКА ОДНОРОДНОСТИ СРЕДНИХ АРИФМЕТИЧЕСКИХ

На практике часто встречаются ситуации, когда среднее значение данных одного эксперимента отличается от среднего значения данных другого, хотя условия эксперимента являются схожими. Тогда возникает вопрос, можно ли считать это расхождение незначимым, т.е. чисто случайным, или оно вызвано существенным различием двух генеральных совокупностей. Например, такие вопросы возникают при исследовании надежности технических систем, где результаты сравниваются с предыдущими измерениями; при контроле качества изделий, изготовленных на разных предприятиях или оборудовании.

Проверка производится с применением t – критерия Стьюдента.

Пусть n_1 и n_2 – объемы выборок; $\overline{x_1}$ и $\overline{x_2}$ – соответствующие средние; s_1^2 и s_2^2 – дисперсии.

1. Оценки дисперсий s_1^2 и s_2^2 однородны. Вычисляется расчетное t – отношение по формуле:

$$t = \frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)}}. \quad (19)$$

Из таблиц распределения Стьюдента при уровне значимости q и числе степеней свободы $f = n_1 + n_2 - 2$ находят табличное значение критерия $t_{таб}$. Если $t_{расч} > t_{таб}$, то расхождение между средними значимо. В противном случае можно принять гипотезу об однородности средних.

2. Оценки дисперсий s_1^2 и s_2^2 неоднородны. Как и в предыдущем случае, здесь можно использовать t – критерий Стьюдента, но формула имеет следующий вид:

$$t = \left|\overline{x_1} - \overline{x_2}\right| \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}. \quad (20)$$

Затем вычисляют величину f по формуле:

$$f = \frac{\left(s_1^2 / n_1 + s_2^2 / n_2\right)^2}{\frac{\left(s_1^2 / n_1\right)^2}{n_1 + 1} + \frac{\left(s_2^2 / n_2\right)^2}{n_2 + 1}} - 2. \quad (21)$$

Найденное значение f округляют до целого и принимают за число степеней свободы. По этой величине и по уровню значимости q из таблиц распределения Стьюдента отыскивается $t_{таб}$. Дальнейший ход проверки не отличается от предыдущего случая.

ЗАДАНИЯ К ЛАБОРАТОРНОЙ РАБОТЕ № 3

1. Определить расчетное значение t – критерия Стьюдента – для выполнения работы «Отбрасывание грубых наблюдений». Проверить, является ли сомнительное наблюдение (подчеркнутое в вариационном ряду) промахом.

2. Определить расчетное значение F – критерия Фишера – и проверить гипотезу об однородности двух дисперсий.

3. Определить расчетное значение G – критерия Кохрена – и проверить гипотезу об однородности нескольких дисперсий.

4. Определить расчетное значение t – критерия Стьюдента – для выполнения работы «Проверка однородности средних». Проверить гипотезу об однородности средних арифметических двух выборок.

ЛАБОРАТОРНАЯ РАБОТА № 4

ПРОВЕРКА НОРМАЛЬНОСТИ РАСПРЕДЕЛЕНИЯ

Эмпирические законы распределения вероятностей имеют дискретный характер независимо от того, является ли эта величина дискретной или непрерывной. Использование такого закона в различных расчётах оказывается не всегда удобным, поэтому возникает задача замены его некоторым теоретическим законом распределения, который был бы в определённом смысле близким к эмпирическому закону. Эта задача решается следующим образом. По виду гистограммы, полигона или графика эмпирической функции распределения по справочникам выбирается подходящий теоретический закон распределения и выдвигается гипотеза о том, что именно этот выбранный закон является истинным законом распределения изучаемой величины. Затем по определённому критерию близости теоретического и эмпирического законов распределений выдвинутая гипотеза принимается или отвергается. Сами критерии близости могут быть различными, поэтому истинность выдвинутой гипотезы можно проверить различным образом.

Одним из наиболее распространённых критериев является критерий χ^2– Пирсона. Вся широта эмпирического распределения разбивается на r частичных интервалов, и сравниваются вероятности попадания измеренных значений случайной величины в эти интервалы для случая эмпирического распределения и для случая теоретического распределения. В качестве меры отклонения берётся величина

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}, \qquad (22)$$

где n – объем выборки;

n_i – число элементов, попавших в i-й интервал;

p_i – вероятность попадания в i-й интервал, вычисленная на основе теоретического распределения.

$$p_i = \Phi(z_2)\text{-}\Phi(z_1)$$
$$z_1 = \frac{x_i^{н} - \bar{x}}{s},\; z_2 = \frac{x_i^{в} - \bar{x}}{s}, \tag{23}$$

где $\bar{x}$ – среднее арифметическое выборки;

s – среднее квадратическое отклонение выборки;

$x_i^{Н}$ – нижняя граница i-го интервала;

$x_i^{В}$ – верхняя граница i-го интервала;

Φ(z) – нормированная функция Лапласа.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-x^{2/2}} dx. \tag{24}$$

Значения ее определяются для $z = z_1$ и $z = z_2$ из приложения 7. При отыскании значений этой функции для отрицательных значений аргумента следует иметь в виду, что функция Φ(z) нечетная: Φ(– z) = – Φ(z).

По выбранному уровню значимости q и числу степеней свободы k = n – 3 из таблиц отыскивается $\chi^2_{таб}$. Гипотезу о нормальности распределения можно принять, если $\chi^2 < \chi^2_{таб}$.

Функция плотности распределения для нормального закона распределения имеет следующий вид:

$$f(x,\mu,s) = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{s}\right)^2\right), \tag{25}$$

где s – среднее квадратическое отклонение;

μ – математическое ожидание случайной или среднее арифметическое величины.

Теоретическая функция распределения выглядит следующим образом:

$$F(x;\mu;s) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{s}\right)^2\right) dx. \tag{26}$$

По определению **эмпирическая функция распределения** – это естественное приближение теоретической функции распределения данной случайной величины, построенное по выборке. По оси абсцисс откладываются интервалы группирования данных, а по оси ординат – накопленная частота.

ЗАДАНИЯ К ЛАБОРАТОРНОЙ РАБОТЕ № 4

1. Определить среднее арифметическое и дисперсию выборки.
2. Рассчитать теоретические вероятности попадания наблюдений в i-й интервал.
3. Вычислить расчетное значение критерия Пирсона. Проверить гипотезу о нормальности распределения вариационного ряда.
4. Построить график функции для нормального теоретического и эмпирического распределения и гистограмму выборки.

ЛАБОРАТОРНАЯ РАБОТА № 5

ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Во многих случаях целью экспериментальных исследований являются установление и изучение зависимости между некоторыми величинами. Если каждая из этих величин является случайной, то используют методы корреляционного анализа.

Будем говорить, что между двумя случайными величинами имеется статистическая связь, если при изменении одной из них меняется распределение другой. Для оценки статистической связи по данным эксперимента широко используется выборочный коэффициент корреляции. Пусть проведено n наблюдений, и в каждом из них определялись значения двух параметров x и y. Найдем по двум выборкам среднее арифметическое $\overline{x}$ и $\overline{y}$, а также среднеквадратическое отклонение s_x, s_y. Выборочный коэффициент корреляции рассчитывается по формуле:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}. \quad (27)$$

Коэффициент корреляции всегда лежит в пределах $-1 \le r_{xy} \le 1$. Он характеризует только линейную зависимость между случайными величинами.

Для определения значимости коэффициента корреляции рассчитывается экспериментальное значение t – критерия Стьюдента:

$$t_{расч} = |r_{xy}|\sqrt{\frac{n-2}{1-r_{xy}^2}}. \quad (28)$$

Его сравнивают с табличным значением t – критерием Стьюдента, найденном при выбранном уровне значимости q и числе степеней свободы

$f = n - 2$. Если $t_{расч} \geq t_{таб}$, то между величинами x и y существует линейная статистическая связь.

Задача линейного регрессионного анализа состоит в восстановлении функциональной зависимости

$$y(x) \equiv M(Y / X = x) = a_0 + a_1 x \tag{29}$$

по результатам измерений

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n). \tag{30}$$

Уравнение (эмпирическая регрессия)

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x \tag{31}$$

определяет прямую, которая является оценкой истинной линии регрессии. Необходимо вычислить точечные и интервальные оценки $\hat{a}_0, \hat{a}_1$ для параметров a_0, a_1 по результатам эксперимента и проверить значимость полученного уравнения регрессии.

Вычисление коэффициентов $\hat{a}_0, \hat{a}_1$ всегда производится с использованием метода наименьших квадратов, но этот метод фиксирует лишь «стратегию» получения эмпирических оценок, допуская различные «тактические приемы», что приводит к большому разнообразию конкретных математических постановок задач, методов и формул получения оценок $\hat{a}_0, \hat{a}_1$ даже в рассматриваемом здесь простейшем случае линейной регрессии. Отметим некоторые из них.

Коэффициенты регрессии можно вычислить:

а) минимизируя сумму квадратов отклонений:

$$E(\hat{a}_0, \hat{a}_1) = \sum_{i=1}^{n} (\hat{a}_0 + \hat{a}_1 x_i - y_i)^2 ; \tag{32}$$

б) численно решая систему уравнений:

$$\frac{\partial E(\hat{a}_0, \hat{a}_1)}{\partial \hat{a}_0} = 0, \qquad \frac{\partial E(\hat{a}_0, \hat{a}_1)}{\partial \hat{a}_1} = 0; \tag{33}$$

в) решая (с использованием точных или итерационных методов) систему нормальных уравнений, предварительно сформировав ее в явном виде:

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}; \tag{34}$$

г) решая систему нормальных уравнений аналитически:

$$\hat{a}_0 = \frac{\sum\limits_{i=1}^{n} y_i \sum\limits_{i=1}^{n} (x_i)^2 - \sum\limits_{i=1}^{n} (x_i y_i) \sum\limits_{i=1}^{n} x_i}{n \sum\limits_{i=1}^{n} (x_i)^2 - \left(\sum\limits_{i=1}^{n} x_i \right)^2}, \hat{a}_1 = \frac{n \sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{n \sum\limits_{i=1}^{n} (x_i)^2 - \left(\sum\limits_{i=1}^{n} x_i \right)^2}, \qquad (35)$$

или

$$\hat{a}_1 = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum\limits_{i=1}^{n} x_i^2 - n\bar{x}^2},$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

или, если предварительно вычислены оценки дисперсий s_X^2, s_Y^2 и коэффициента корреляции $\hat{\rho}$:

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad s_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2, \qquad (36)$$

$$r_{xy} = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2 \sum\limits_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{1}{(n-1) s_X s_Y} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}), \qquad (37)$$

то

$$\hat{a}_1 = r_{xy} \frac{s_Y}{s_X}, \quad \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}.$$

Доверительные интервалы для коэффициентов регрессии $\hat{a}_0, \hat{a}_1$, соответствующие доверительной вероятности $p = 1 - \alpha$, имеют вид:

$$\hat{a}_0 - t_{\alpha,n-2} s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}} < a_0 < \hat{a}_0 + t_{\alpha,n-2} s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}},$$

$$\hat{a}_1 - \frac{t_{\alpha,n-2} s}{\sqrt{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}} < a_1 < \hat{a}_1 + \frac{t_{\alpha,n-2} s}{\sqrt{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}} \qquad (38)$$

или

$$\hat{a}_0 - t_{\alpha,n-2}\sqrt{\left(\frac{1}{n}+\frac{\bar{x}^2}{(n-1)s_X^2}\right)\frac{n-1}{n-2}s_Y^2\left(1-\hat{\rho}\right)} < a_0 <$$

$$< \hat{a}_0 + t_{\alpha,n-2}\sqrt{\left(\frac{1}{n}+\frac{\bar{x}^2}{(n-1)s_X^2}\right)\frac{n-1}{n-2}s_Y^2\left(1-\hat{\rho}\right)}, \qquad (39)$$

$$\hat{a}_1 - t_{\alpha,n-2}\frac{\;}{s_X\sqrt{n-2}} < a_1 < \hat{a}_1 + t_{\alpha,n-2}\frac{\;}{s_X\sqrt{n-2}},$$

где $t_{\alpha,n-2}$ - квантиль распределения Стьюдента, определяемый как корень уравнения:

$$F_{n-2}\left(t_{\alpha,n-2}\right)=1-\alpha/2,$$

где $F_{n-2}(t)$ - функция распределения Стьюдента с $n-2$ степенями свободы.

Доверительная область для всей линии регрессии определяется с помощью уравнений:

$$y'(x)=\hat{a}_0+\hat{a}_1x-s\sqrt{2f_{\alpha,2,n-2}}\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}},$$

$$y''(x)=\hat{a}_0+\hat{a}_1x+s\sqrt{2f_{\alpha,2,n-2}}\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}} \qquad (40)$$

описывающих соответственно нижнюю и верхнюю границы области («полосы»), в которой с доверительной вероятностью $p=1-\alpha$ лежит истинная линия регрессии. Здесь $f_{\alpha,2,n-2}$ - квантиль распределения Фишера, определяемый как решение уравнения:

$$F_{2,n-2}\left(f_{\alpha,2,n-2}\right)=1-\alpha; \qquad (41)$$

где $F_{2,n-2}(x)$ – функция распределения Фишера с 2 и $n-2$ степенями свободы,

s^2 – «остаточная» дисперсия, характеризующая рассеяние экспериментальных точек относительно линии регрессии

$$s^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2. \qquad (42)$$

Для проверки значимости уравнения регрессии в целом используется критерий Фишера: если

$$\frac{s_Y^2}{s^2} > f_{\alpha, n-1, n-2}, \tag{43}$$

то уравнение регрессии адекватно (статистически значимо) описывает результаты эксперимента при $(100 \cdot \alpha)$ - процентном уровне значимости.

Отношение (полной и остаточной дисперсий) s_Y^2 / s^2 показывает, во сколько раз уравнение регрессии предсказывает результаты опыта лучше, чем среднее $\bar{y}$. Необходимо помнить, что доверительная оценка отклонения эмпирической линии регрессии от теоретической существенно ухудшается по мере удаления от среднего значения $\bar{x}$. В частности, по этой причине опасна экстраполяция эмпирической регрессионной зависимости за пределы интервала (x_1, x_n), для которого она получена.

ЗАДАНИЯ К ЛАБОРАТОРНОЙ РАБОТЕ № 5

1. Вычисление коэффициентов линейной регрессионной зависимости и статистический анализ полученного уравнения.
2. Для уровня значимости q вычислить доверительные интервалы для коэффициентов регрессии.
3. Изобразите на одном графике линию регрессии и границы доверительной области для нее.
4. Проверьте адекватность полученного уравнения регрессии по критерию Фишера.

ЛАБОРАТОРНАЯ РАБОТА № 6

ПРОВЕДЕНИЕ МНОГОФАКТОРНОГО ЭКСПЕРИМЕНТА

После изучения объекта исследования и его физической сущности возникает ряд представлений о действии различных параметров (факторов) и необходимость получить экспериментальные данные о их совокупном влиянии на какой-либо показатель (критерий), характеризующий объект исследований. При составлении плана эксперимента прежде всего назначают (выбирают) независимые факторы, исходя из априорной (доопытной) информации или предварительного изучения объекта исследования.

Факторы бывают количественные и качественные. Количественные можно измерить и выразить в числах.

Далее при составлении плана эксперимента назначаются уровни варьирования факторами, или градации.

В многофакторных экспериментах, основанных на дисперсионном анализе, обычно берут два уровня факторов (верхний и нижний). Ведь в таких экспериментах важно проверить, значимо ли влияет тот или иной фактор и есть ли факторные взаимодействия. Варьирование переменными на двух уровнях позволяет значительно уменьшить объем экспериментальной и счетной работы.

В многофакторном эксперименте уровни одного фактора должны сочетаться с уровнями другого, образуя тем самым вариант испытаний. Интервал варьирования того или иного фактора должен быть таким, чтобы можно было реализовать любой вариант испытаний.

В теории планирования эксперимента условно принято обозначать нижнюю границу или нижний уровень фактора знаком - 1, а верхнюю – знаком +1.

В теории планирования эксперимента, если факторы устанавливаются на двух уровнях, принято обозначать – эксперимент типа 2^n, где n – число факторов. Если факторы устанавливаются на трех уровнях, то называют эксперимент типа 3^n и т.д.

Комбинации условий эксперимента 2^2 можно выразить в виде таблицы, если обозначить нижний уровень фактора -1, а верхний +1. Такая таблица называется матрицей планирования эксперимента (табл. 1).

Таблица 1

Матрица планирования двухфакторного эксперимента

№ опыта	Факторы и их взаимодействия			Кодовое обозначение
	A	B	AB	
1	2	3	4	5
1	-	-	+	(1)
2	+	-	-	a
3	-	+	-	b
4	+	+	+	ab

Примечание. С целью упрощения записи символы 1 не указаны, а проставлены только их знаки.

В первом столбце рассматриваемой матрицы эксперимента записаны номера опытов (без повторений), которые при реализации необходимо рандомизировать. Это варианты испытаний. Второй и третий столбцы представляют собственно планирование, образуя возможные комбинации знаков факторов (условий испытаний). Четвертый столбец образуется перемножением знаков факторов A×B. Он показывает возможные взаимодействия факторов и их знаки в том или ином опыте.

Взаимодействие факторов в многофакторном эксперименте следует понимать так: изменение одного фактора сопровождается непропорциональными изменениями результатов эксперимента при изменении уровней другого.

Рассмотрим планирование экспериментов с тремя независимыми переменными (факторами) A, B, C, которые также варьируют на двух уровнях. Это будет эксперимент типа 2^3.

Однако при постановке полного многофакторного эксперимента даже только на двух уровнях уже требуется проводить большое число опытов, поэтому для сокращения затрат средств и времени можно не реализовать полный факторный эксперимент, а ограничиться лишь некоторой частью его. В этих случаях используют часть матрицы, называемую дробной репликой (табл. 2).

Таблица 2

Матрица планирования трехфакторного эксперимента

№ опыта	Факторы и их взаимодействия							Кодовое обозначение строк
	A	B	C	AB	AC	BC	ABC	
1	-	-	-	+	+	+	-	(1)
2	+	-	-	-	-	+	+	a
3	-	+	-	-	+	-	+	b
4	+	+	-	+	-	-	-	ab
5	-	-	+	+	-	-	+	c
6	+	-	+	-	+	-	-	ac
7	-	+	+	-	-	+	-	bc
8	+	+	+	+	+	+	+	abc

При реализации полного факторного эксперимента можно определить взаимодействия всех порядков, но в этом нет необходимости. При реализации же дробной реплики от полного факторного эксперимента теряется часть информации о влиянии факторов, а именно в дробных репликах эффекты ряда взаимодействий (в зависимости от структуры дробной реплики) приравниваются либо какому-нибудь эффекту фактора, либо ошибке эксперимента. Или иначе – эффект фактора смешивается с взаимодействием высокого порядка. Если известно заранее, что взаимодействие, с которым смешан фактор, не значимо, то вычисленный эффект будет равен эффекту фактора.

Дробная реплика может быть построена с различной степенью дробности. Допустим, что требуется получить полуреплику от трехфакторного эксперимента: a, b, ab, c, ac, bc, abc. Из восьми условий эксперимента нужно получить четыре. Факторы обозначим через x_1, x_2, x_3. Если известно, что взаимодействие x_1, x_2 не значимо, то можно с ним смешать фактор x_3 и можно записать $x_3 = x_1\ x_2$. Выражение $x_3 = x_1\ x_2$ называется генерирующим соотношением. При помощи генерирующих соотношений строятся дробные реплики.

В трехфакторном эксперименте генерирующие соотношения можно принять в виде: $x_3 = -\ x_1\ x_2$, $x_2 = x_1\ x_3$, $x_2 = -\ x_1\ x_3$ и т.д.

Задавшись определяющим контрастом, можно найти генерирующие соотношения, которые покажут, какие эффекты в эксперименте будут смешаны.

Для трехфакторного эксперимента определяющие контрасты можно задать так:

$$I = x_1\ x_2\ x_3\ ,\ I = -\ x_1\ x_2\ x_3. \qquad (44)$$

Определяющий контраст I всегда при кодированных факторах (+1 или -1) будет равен единице.

Для того, чтобы найти генерирующие соотношения, или, как еще говорят, определить совместимые оценки эффектов факторов, необходимо последовательно помножить независимые переменные на определяющий контраст, учитывая $x_1^2 = 1 = I$.

Поясним подробнее, почему так получается. Задавшись определяющим контрастом. $I = x_1 x_2 x_3$, можно получить одну полуреплику от трехфакторного эксперимента. Допустим, для анализа выбрали полуреплику с нечетным сочетанием символов a, b, c, abc. Составим для этой полуреплики развернутую матрицу планирования эксперимента (табл. 3).

Таблица 3

Полуреплика факторного эксперимента 2^3

Кодовое обозначение строк (комбинаций условий)	Эффекты факторов						
	x_1	x_2	x_1x_2	x_3	x_1x_3	x_2x_3	$x_1x_2x_3$
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
c	-	-	+	+	-	-	+
abc	+	+	+	+	+	+	+

Из таблицы 3 видно, что эффект фактора x_1 :

$$x_1= +a-b-c+abc. \tag{45}$$

Эффект взаимодействия $x_2 x_3$:

$$x_2 x_3= +a-b-c+abc. \tag{46}$$

Значит, при реализации этой полуреплики нельзя различить эффекты факторов x_1 и x_2x_3, т.е. они будут совместными. Из таблицы 3 видны и другие совместные эффекты.

В этом плане нельзя различить отдельно эффект фактора и парных взаимодействий при наличии эффектов взаимодействий.

Таким образом, если экспериментатор не знает заранее, что взаимодействий нет, то реализация такого плана не имеет практической ценности.

Как показывает опыт экспериментальной работы, полуреплики от полного факторного эксперимента целесообразно получать лишь при числе факторов более трех.

Вышеизложенное о разрешающей способности дробных реплик и правилах их построения можно свести в общую схему. Для этого необходимо проделать перечисленные ниже операции.

1. Задавшись общим количеством факторов n, необходимо выписать общее количество условий испытаний или опытов N по формуле $N=2^n$, если факторы варьируют на двух уровнях.

2. Исходя из конкретной обстановки, условий, бюджета времени и т.п., решается вопрос об объеме реализации опытов. Если эксперимент решено провести не в полном объеме, то назначается дробная реплика (1/2, 1/8 и т.д.).

3. Задавшись дробной репликой, вычисляют величину p в формуле $Q=2^{n-p}$, где Q – число опытов в дробной реплике.

Это необходимо для того, чтобы знать, сколько эффектов факторов смешать с эффектами взаимодействий.

4. Решается вопрос о том, какие факторы и с какими взаимодействиями можно смешать.

Если исследование проводится впервые, то делают случайный выбор и на этом основании строят генерирующие отношения. Для удобства построения дробной реплики эффект факторов лучше смешивать с эффектами взаимодействий, которые содержат нижние индексы их начальных цифр, например $x_2 = x_1 x_2 x_3$ или $x_3 = x_1 x_2 x_3 x_4 x_5$ и т.п.

5. Строится матрица планирования эксперимента дробной реплики на основании выбранных генерирующих соотношений.

6. При выбранных генерирующих соотношениях записываются обобщающие контрасты и затем определяется обобщающий определяющий контраст.

7. Обобщающий определяющий контраст умножается на тот или иной фактор (взаимодействие) для определения совместных эффектов факторов (взаимодействий). Как правило, взаимодействиями высоких порядков, начиная с тройных, пренебрегают.

Математический анализ результатов эксперимента проводится в следующей последовательности:

1. Вычисление средних построчных значений выходного параметра:

$$\bar{y} = \frac{1}{\gamma} y_{ul}, \tag{47}$$

где γ – число повторений опыта.

2. Вычисление построчных дисперсий:

$$S_u^2 = \frac{1}{\gamma - 1} \sum_{l=1}^{\gamma} (y_{ul} - \bar{y}_u)^2. \tag{48}$$

3. Проверка однородности наблюдений. Гипотеза об однородности выборочных оценок S^2 $\{y_u\}$ не отвергается, если:

$$\frac{S_{max}^2}{\sum\limits_{u=1}^{N} S_u^2} \le G_\alpha(\gamma - 1; N), \tag{49}$$

где S^2_{max} – наибольшее из вычисленных по (48) значение построчных дисперсий;

$G_\alpha(\gamma - 1, N)$ – табличное значение критерия Кохрена;

q = 0,05 – уровень значимости;

N – количество опытов в плане.

4. Оценка дисперсии воспроизводимости единичного опыта:

$$S^2\{y\} = \frac{1}{N}\sum_{u=1}^{N} S_u^2 ; \tag{50}$$

с числом степеней свободы: $f_1 = N(\gamma - 1)$.

5. Оценка дисперсии воспроизводимости среднего значения из γ наблюдений:

$$S^2\{\bar{y}\} = \frac{1}{\gamma} S^2\{y\}; \tag{51}$$

с числом степеней свободы также f_1.

6. Определение коэффициентов регрессии производится по следующим формулам:

- свободный член:

$$в_0 = \frac{1}{N}\sum_{u=1}^{N} \bar{y}_u ; \tag{52}$$

- коэффициенты при линейных членах:

$$в_i = \frac{1}{N}\sum_{u=1}^{N} x_{iu}\bar{y}_u ; \quad i = 1, 2 \ldots m, \tag{53}$$

где m – число независимых переменных;

- коэффициенты при взаимодействиях;

$$\begin{gathered} в_{ij} = \frac{1}{N}\sum_{u=1}^{N} x_{iu}x_{ju}\bar{y}_u \;\; i, j = 1, 2 \ldots m; i \ne j; \\ в_{ijq} = \frac{1}{N}\sum_{u=1}^{N} x_{iu}x_{ju}x_{qu}\bar{y}_u \;\; i, j, q = 1, 2 \ldots m; i \ne j \ne q. \end{gathered} \tag{54}$$

7. Оценки дисперсии коэффициентов регрессии:

$$S^2\{в_i\} = \frac{1}{\gamma N} S^2\{y\}. \qquad (55)$$

8. Проверка статистической значимости коэффициентов регрессии. Коэффициент регрессии не значим, если:

$$|в_i| \langle S\{в_i\} \cdot t_\alpha(f_1), \qquad (56)$$

где $t_\alpha(f_1)$ – табличное значение критерия Стьюдента при уровне значимости q = 0,05 и числе степеней свободы $f_1 = N(\gamma - 1)$.

Коэффициенты, значения которых удовлетворяют условию (56) исключают соответствующие члены из уравнения регрессии.

9. Оценка дисперсии для проверки адекватности:

$$S_{ад}^2 = \gamma \sum_{u=1}^{N} (\bar{y}_u - \hat{y}_u)^2, \qquad (57)$$

где $\hat{y}_u$ – вычисленное по уравнению регрессии значение выходного параметра в u - м опыте.

Число степеней свободы

$$F_{ад} = N - K, \qquad (58)$$

где K – общее число членов в уравнении регрессии.

10. Проверка адекватности модели. Гипотеза об адекватности представления поверхности отклика уравнением регрессии, содержащим K членов, не отвергается, если:

$$F_{ад} = \frac{S_{ад}^2}{F_{ад} S^2\{y\}} \le F_\alpha(f_2; f_1), \qquad (59)$$

где $F_\alpha(F_{ад}; f_1)$ – значение критерия Фишера при уровне значимости q = 0,05; и числе степеней свободы $F_{ад}$ и f_1.

ЗАДАНИЯ К ЛАБОРАТОРНОЙ РАБОТЕ № 6

1. Определить область планирования эксперимента, число действующих факторов, функцию отклика.

2. Провести проверку экспериментальных данных на однородность и нормальность.

3. Получить уравнение регрессии. Провести сравнение экспериментальных и расчетных значений. Результаты занести в отчет.

4. Провести оценку значимости коэффициентов регрессии и оценку адекватности полученного уравнения. Результаты занести в отчет.

5. Провести анализ типа поверхности отклика, построить линии равного уровня. Результаты занести в отчет. Схематично изобразить в отчете полученную поверхность, линии равного уровня.

6. Получить уравнение регрессии. Провести сравнение экспериментальных и расчетных значений. Занести результаты в отчет.

7. Провести оценку значимости коэффициентов регрессии и оценку адекватности полученного уравнения. Результаты занести в отчет.

ПРИЛОЖЕНИЯ

Приложение 1

Математическая обработка результатов эксперимента в Excel

1. СРЗНАЧ. – определение среднего значения.
2. ДИСП (число1;число2) – оценивает дисперсию по выборке.
3. СРОТКЛ – определяет среднее абсолютных значений отклонений точек данных от среднего.
4. СТАНДОТКЛОН – оценивает среднее квадратическое отклонение по выборке.
5. КВАДРОТКЛ – оценивает сумму квадратов отклонений точек данного от среднего по выборке.
6. ДОВЕРИТ(альфа; стандартное откл; размер) – определяет доверительный интервал для среднего генеральной совокупности, где:

   а) альфа – уровень значимости, используемый для вычисления уровня надежности экспериментальных данных (a = 1 - P);

   б) стандартное откл. – среднее отклонение генеральной совокупности для интервала данных (предполагается известным, т.е. ранее рассчитывается);

   в) размер – размер выборки (f = n – 1, где n – число измерений).
7. СТЬЮДРАСПОБР(альфа; размер) – определяет значение критерия Стьюдента.
8. СКОС(число1;число2) – определяет коэффициент асимметрии по выборке.
9. ЭКСЦЕСС(число1;число2) – определяет коэффициент эксцесса по выборке.
10. FРАСПОБР(q; $f_1;f_2$) - определение критерия Фишера.
11. ХИ2ОБР(q; k) - определение χ^2 критерия Пирсона.
12. КОРРЕЛ(массив1; массив2) - определение коэффициента корреляции:

    Массив 1 – первый интервал ячеек со значениями;

    Массив 2 – второй интервал ячеек со значениями.
13. НОРМСТРАСП (число) - расчет функции Лапласа.

Приложение 2

Математическая обработка результатов эксперимента в MATHCAD

1. mean(A) – функция вычисляет значение выборочного среднего.
2. $\frac{nm}{nm-1}var(A)$ - определяет дисперсию.
3. stdev(x) – среднее квадратическое отклонение;
4. $qt\left(1-\frac{q}{2}, nm-1\right)$ – определение критерия Стьюдента при уровне значимости q.
5. Histogram(n, X) – построение гистограммы, где: n – число интервалов, на которое разбивается весь диапазон исходных данных X. Эта функция возвращает 2 столбца. В первом содержатся средние точки каждого из n интервалов, во втором – частоты попадания исходных данных X в каждый из n интервалов.
6. Hist(int, X) - построение гистограммы, где int либо вектор середин интервалов (можно задать интервалы разной ширины), либо число интервалов.
7. skew(A) – определяет коэффициент асимметрии по выборке A;
8. kurt(A) – определяет коэффициент эксцесса по выборке A.
9. corr(X, Y) - определение коэффициента корреляции, где:
   X – первый массив данных со значениями;
   Y – второй массив данных со значениями.
10. intersept(X, Y) – коэффициент b_0 в уравнении линейной регрессии.
11. slope(X, Y) – коэффициент b_1 в уравнении линейной регрессии.
12. qchisq(1 – q, k) – определение χ^2 критерия Пирсона.
13. qF(1– q; $f_1;f_2$) - определение критерия Фишера.

Приложение 3

Значения критерия Стьюдента при различной доверительной вероятности (P) для разного числа измерений (n)

n	P								
	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
2	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66	636,62
3	0,82	1,06	1,39	1,89	2,92	4,30	6,96	9,92	31,60
4	0,76	0,98	1,25	1,64	2,35	3,18	4,54	5,84	12,92
5	0,74	0,94	1,19	1,53	2,13	2,78	3,75	4,60	8,61
6	0,73	0,92	1,16	1,48	2,02	2,57	3,36	4,03	6,87
7	0,72	0,91	1,13	1,44	1,94	2,45	3,14	3,71	5,96
8	0,71	0,90	1,12	1,41	1,89	2,36	3,00	3,50	5,41
9	0,71	0,89	1,11	1,40	1,86	2,31	2,90	3,36	5,04
10	0,70	0,88	1,10	1,38	1,83	2,26	2,82	3,25	4,78
11	0,70	0,88	1,09	1,37	1,81	2,23	2,76	3,17	4,59
12	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11	4,44
13	0,70	0,87	1,08	1,36	1,78	2,18	2,68	3,05	4,32
14	0,69	0,87	1,08	1,35	1,77	2,16	2,65	3,01	4,22
15	0,69	0,87	1,08	1,35	1,76	2,14	2,62	2,98	4,14
16	0,69	0,87	1,07	1,34	1,75	2,13	2,60	2,95	4,07
17	0,69	0,86	1,07	1,34	1,75	2,12	2,58	2,92	4,01
18	0,69	0,86	1,07	1,33	1,74	2,11	2,57	2,90	3,97
19	0,69	0,86	1,07	1,33	1,73	2,10	2,55	2,88	3,92
20	0,69	0,86	1,07	1,33	1,73	2,09	2,54	2,86	3,88
30	0,68	0,85	1,06	1,31	1,70	2,05	2,46	2,76	3,66
40	0,68	0,85	1,05	1,30	1,68	2,02	2,43	2,71	3,56
60	0,68	0,85	1,05	1,30	1,67	2,00	2,39	2,66	3,46

Приложение 4

Значения G – критерия Кохрена – для уровня значимости α = 0,05 в зависимости от числа степеней свободы m_1 и m_2

m_2	m_i												
	1	2	3	4	5	6	7	8	9	10	16	36	144
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8333	0,8159	0,8010	0,7880	0,7341	0,6602	0,5813
3	0,9669	0,8709	0,7977	0,7457	0,7071	0,6771	0,6530	0,6333	0,6167	0,6025	0,5466	0,4748	0,4031
4	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017	0,4884	0,4366	0,3720	0,3093
5	0,8412	0,6838	0,5981	0,5441	0,5065	0,4783	0,4564	0,4387	0,4241	0,4118	0,3645	0,3066	0,2513
6	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682	0,3566	0,3135	0,2612	0,2919
7	0,7271	0,5612	0,4800	0,4307	0,3974	0,3726	0,3535	0,3384	0,3259	0,3154	0,2756	0,2278	0,1833
8	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926	0,2829	0,2462	0,2022	0,1616
9	0,6385	0,4775	0,4027	0,3584	0,3286	0,3067	0,2901	0,2768	0,2659	0,2568	0,2226	0,1820	0,1446
10	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,2666	0,2541	0,2439	0,2355	0,2032	0,1655	0,1308
12	0,5410	0,3924	0,3264	0,2880	0,2626	0,2439	0,2299	0,2187	0,2098	0,2020	0,1737	0,1403	0,1100
15	0,4709	0,3346	0,2758	0,2419	0,2195	0,2034	0,1911	0,1815	0,1736	0,1671	0,1429	0,1144	0,0889
20	0,3894	0,2705	0,2205	0,1921	0,1735	0,1602	0,1501	0,1422	0,1357	0,1303	0,1108	0,0879	0,0675
24	0,3434	0,2354	0,1907	0,1656	0,1493	0,1374	0,1286	0,1216	0,1160	0,1113	0,0942	0,0743	0,0567
30	0,2929	0,1980	0,1593	0,1377	0,1237	0,1137	0,1061	0,1002	0,0958	0,0921	0,0771	0,0604	0,0457
40	0,2370	0,1576	0,1259	0,1082	0,0968	0,0887	0,0827	0,0780	0,0745	0,0713	0,0595	0,0462	0,0347
60	0,1737	0,1131	0,0895	0,0765	0,0682	0,0623	0,0583	0,0552	0,0520	0,0497	0,0711	0,0316	0,0245
120	0,0998	0,0632	0,0495	0,0419	0,0371	0,0337	0,0312	0,0292	0,0279	0,0266	0,0218	0,0168	0,0120

Приложение 5

Значения критерия Фишера (F – критерия) при доверительной вероятности 0,95 и числа степеней свободы m_1 и m_2

m_2	m_1															
	1	2	3	4	5	6	7	8	9	10	12	15	30	40	60	120
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91	245,95	250,10	251,14	252,20	253,25
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,46	19,47	19,48	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,62	8,59	8,57	8,55
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,75	5,72	5,69	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,50	4,46	4,43	4,40
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,81	3,77	3,74	3,70
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,38	3,34	3,30	3,27
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,08	3,04	3,01	2,97
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,86	2,83	2,79	2,75
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,70	2,66	2,62	2,58
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,57	2,53	2,49	2,45
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,47	2,43	2,38	2,34
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,38	2,34	2,30	2,25
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,31	2,27	2,22	2,18
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,25	2,20	2,16	2,11
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,19	2,15	2,11	2,06
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,15	2,10	2,06	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,11	2,06	2,02	1,97
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,07	2,03	1,98	1,93
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,04	1,99	1,95	1,90
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,01	1,96	1,92	1,87
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	1,98	1,94	1,89	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	1,96	1,91	1,86	1,81
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	1,94	1,89	1,84	1,79
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	1,92	1,87	1,82	1,77
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,90	1,85	1,80	1,75
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,88	1,84	1,79	1,73
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,87	1,82	1,77	1,71
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,85	1,81	1,75	1,70
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,84	1,79	1,74	1,68
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,74	1,69	1,64	1,58
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,65	1,59	1,53	1,47
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,55	1,50	1,43	1,35

Приложение 6

Значения χ^2 – критерия Пирсона – для уровня значимости α в зависимости от числа степеней свободы m

m	A										
	0,995	0,99	0,975	0,95	0,9	0,5	0,1	0,05	0,025	0,01	0,005
1	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	12,340	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	21,337	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672
40	20,707	22,164	24,433	26,509	29,051	39,335	51,805	55,758	59,342	63,691	66,766
50	27,991	29,707	32,357	34,764	37,689	49,335	63,167	67,505	71,420	76,154	79,490
60	35,534	37,485	40,482	43,188	46,459	59,335	74,397	79,082	83,298	88,379	91,952
70	43,275	45,442	48,758	51,739	55,329	69,334	85,527	90,531	95,023	100,425	104,215
80	51,172	53,540	57,153	60,391	64,278	79,334	96,578	101,879	106,629	112,329	116,321
90	59,196	61,754	65,647	69,126	73,291	89,334	107,565	113,145	118,136	124,116	128,299
100	67,328	70,065	74,222	77,929	82,358	99,334	118,498	124,342	129,561	135,807	140,169

СПИСОК ЛИТЕРАТУРЫ

1. Адлер, Ю.П. Планирование эксперимента при поиске оптимальных условий / Ю.П. Адлер, Е.В. Маркова, Ю.В. Грановский. – М.: Наука, 1976. – 280 с.

2. Гмурман, В.Е. Теория вероятности и математическая статистика / В.Е. Гмурман. – М.: Высшая школа, 1998. – 479 с.

3. Ивановский, Р.И. Теория вероятностей и математическая статистика. Основы, прикладные аспекты с применением и задачами в среде Mathcad / Р.И. Ивановский. – СПб: БХВ – Петербург, 2008. – 528 с.

4. Пен, Р.З. Статические методы моделирования и оптимизации процессов ЦБП / Р.З. Пен. – Красноярск: Изд-во Красноярского университета, 1982. – 192 с.

5. Пижурин, А.А. Исследования процессов деревообработки / А.А. Пижурин, М.С. Розенблит. – М.: Лесная промышленность, 1984. – 231 с.

СОДЕРЖАНИЕ

Введение ……………………………………………………………………3

Лабораторная работа № 1. Исследование статистических характеристик случайных величин ………………………………………………………………4

Задания к лабораторной работе № 1……………………………...…...……5

Лабораторная работа № 2. Метод группирования данных……………..………6

Задания к лабораторной работе № 2…………………………………...……10

Лабораторная работа № 3. Проверка статистических гипотез……………..…...10

3.1. Отбрасывание грубых наблюдений……………………………….……….10

3.2. Проверка гипотезы об однородности двух дисперсий………….……….…11

3.3. Проверка однородности нескольких дисперсий, найденных по выборкам одинакового объема……………………………………………………….......11

3.4. Проверка однородности нескольких дисперсий, найденных по выборкам различного объема…………………………………………...……12

3.5. Проверка однородности средних арифметических………………………12

Задания к лабораторной работе № 3……………………………………….....14

Лабораторная работа № 4. Проверка нормальности распределения……......…14

Задания к лабораторной работе № 4……………………………………..…......16

Лабораторная работа № 5. Определение коэффициента корреляции…..……...16

Задания к лабораторной работе № 5…………………………………...…….20

Лабораторная работа № 6. Проведение многофакторного эксперимента…..….20

Задания к лабораторной работе № 6………………………………………..….26

Приложения……………………………………………………………...…....28

Список литературы……..………...………………………………...……34